Journal of Statistical Software

February 2009, Volume 29, Book Review 13. http://www.jstatsoft.org/

Reviewer: Patrick Mair
Wirtschaftsuniversität Wien

Analysis of Multivariate Social Science Data (2nd Edition)

David J. Bartholomew, Fiona Steele, Irini Moustaki, Jane I. Galbraith
Chapman & Hall/CRC, Boca Raton, FL, 2008.
ISBN 978-1-58488-960-1. 371 pp. USD 69.95 (P).
http://www.cmm.bristol.ac.uk/team/amssd.shtml

When four of the leading researchers in the field of quantitative social sciences team up to write a book together, you can expect nothing less than a brilliant work. That is what the first edition of "Analysis of Multivariate Social Science Data" from 2002 was, and that's what the current second edition is. This new edition contains additional chapters on regression analysis, confirmatory factor analysis including structural equation models, and multilevel models.

The strength of this book lies in the right mixture of simple mathematical expressions, comprehensive non-mathematical descriptions of various multivariate approaches, numerous interesting real-life data examples (almost half of each chapter is dedicated to examples), and, last but not least, detailed interpretation of the results. As in other Bartholomew books, well-known methods or certain parts of them are, in many cases, presented from a slightly different angle. This makes this book also interesting for experienced researchers, as certain explanations they give often add a new perspective to a particular statistical method.

The comprehensive Web resource the authors provide is also commendable. The data sets used in this book can be downloaded from there, and SPSS syntax and output files with the results of the corresponding analysis are online as well. For the methods of Chapters 8–11 (see below) they offer the **LAMI** software, an interface to **GENLAT** and **LATCLASS**. One would have liked to see some R code included, since all the methods covered in this book are available in various R packages and R is becoming more and more popular in social sciences and related applied fields.

What follows is a short description of each of the 12 chapters of the book. Chapter 1 describes the general concept of the book, Chapter 2 deals with the first statistical method: cluster analysis, or more precisely, various approaches for hierarchical clustering such as single linkage and complete linkage.

Multidimensional scaling (MDS) is described in the following chapter. The authors focus on metric as well as on non-metric MDS, and include also a section about model fit. Numerous examples are provided to show various kinds of graphical representations of the results. A

method that is somewhat related to MDS is correspondence analysis (CA) which is described in Chapter 4. The authors introduce CA in "Greenacre manner", with basic terms such as profiles, masses, inertia, symmetric and non-symmetric biplots. Again, the examples in the second part of the chapter give several insights into the application and interpretation of CA. The last chapter of dimension-reducing methods deals with principal components analysis (PCA). A very interesting section addresses the link between PCA, CA, and MDS which, I think, is very important for the general understanding of the whole concept of scaling.

One of the new chapters in this second edition is Chapter 6, "Regression Analysis". With respect to a deeper understanding of the subsequent methods related to the linear model formulation, it is important to have such a general chapter in the book. The authors include a section about path models which links directly to structural equation models (SEM) described later.

The next five chapters are about latent variable models. Chapter 7 introduces "classical" factor analysis. It explains the linear model formulations, factor loadings, communalities, and provides a section on goodness-of-fit/number of factors. Basic rotation techniques are explained as well as demonstrated on various real-life examples.

The following two chapters are very important for applications in social sciences, since researchers have to deal commonly with categorical data. Chapter 8 is about binary factor analysis. It embeds item response theory (IRT) into this context. The underlying variable approach is presented in a very accessible manner as well. Chapter 9 presents ordinal factor analysis in an analogous manner (i.e., polytomous IRT, underlying variable approach).

Chapter 10 covers latent class analysis (LCA) for parametric clustering of binary data. This method does not really belong to the standard repertoire of multivariate statistical methods as covered in other textbooks. All the better that it is elaborated in this book since LCA is a very useful tool for analyzing binary data. Also appreciated is that the authors link LCA to IRT on the one hand, and to cluster analysis on the other.

Incorporating such a comprehensive topic as SEM as a single chapter into a textbook on multivariate analysis is a difficult task. Nevertheless, the authors did a great job in presenting concisely the most important tasks, such as the measurement model (confirmatory factor analysis), the structural model, model identification, model fit, and model interpretation by means of simple examples.

An additional issue that occurs quite often in the social sciences is that corresponding populations have a hierarchical structure. This setting can be statistically approached by means of multilevel models which are described in Chapter 12. The concept of fixed vs. random effects is explained as well as the corresponding estimation of group effects. These concepts are extended in terms of multilevel multivariate regression models and multilevel factor models.

Overall, this is an outstanding book on multivariate statistics in the field of social sciences, with a strong focus on categorical data. It can be recommended without reservations for quantitative graduate courses in psychology, sociology, education and related areas. Only very basic, introductory statistical knowledge is required to read and understand this book. In addition, it provides a solid base for gaining access into other prominent, more advanced latent variable books such as Skrondal and Rabe-Hesketh (2004) and Bartholomew and Knott (1999).

If a third edition were under consideration, it would be useful to have a chapter on optimal scaling (see Gifi 1990, as a comprehensive reference). In the current edition this approach is

just mentioned briefly in the CA chapter, but I think that these models solve many practical problems whenever researchers have to deal with categorical or mixed scale levels.

References

Bartholomew DJ, Knott M (1999). *Latent Variable Models and Factor Analysis*. Number 7 in Kendall's Library of Statistics, 2nd edition. Hodder Arnold, London.

Gifi A (1990). *Nonlinear Multivariate Analysis*. John Wiley & Sons, Chichester, England.

Skrondal A, Rabe-Hesketh S (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC, Boca Raton, FL.

Reviewer:

Patrick Mair
Wirtschaftuniversität Wien
Department of Statistics and Mathematics
Augasse 2–6
1090 Vienna, Austria
E-mail: Patrick.Mair@wu-wien.ac.at
URL: http://statmath.wu-wien.ac.at/~mair/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 29, Book Review 13 *Published:* 2009-02-18
February 2009